Elliptical Distributions

Max Turgeon

STAT 7200–Multivariate Statistics

Density contours i

- Recall the density of the multivariate normal distribution:

$$f(\mathbf{Y}) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{Y} - \mu)^T \Sigma^{-1} (\mathbf{Y} - \mu)\right).$$

- For a real number $k > 0$, we can look at the values of $\mathbf{Y}$ for which $f(\mathbf{Y}) = k$. We have

Density contours ii

$$
\begin{aligned}
f(\mathbf{Y}) = k \Leftrightarrow \frac{1}{\sqrt{(2\pi)^p|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{Y}-\mu)^T\Sigma^{-1}(\mathbf{Y}-\mu)\right) = k \\
\Leftrightarrow \exp\left(-\frac{1}{2}(\mathbf{Y}-\mu)^T\Sigma^{-1}(\mathbf{Y}-\mu)\right) = k\sqrt{(2\pi)^p|\Sigma|} \\
\Leftrightarrow (\mathbf{Y}-\mu)^T\Sigma^{-1}(\mathbf{Y}-\mu) = -2\log\left(k\sqrt{(2\pi)^p|\Sigma|}\right).
\end{aligned}
$$

Density contours iii

- In other words, the sets of constant density correspond to the sets where the quadratic form

$$(\mathbf{Y} - \mu)^T \Sigma^{-1} (\mathbf{Y} - \mu)$$

  is constant.
- The latter sets are **ellipses** (or multivariate generalizations thereof).

Example i

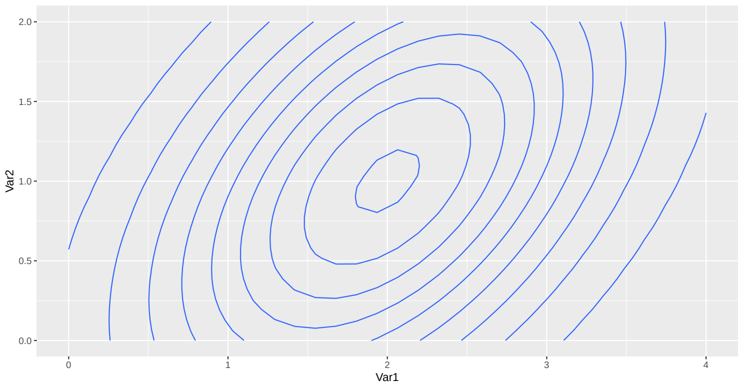
```r
library(mvtnorm)

mu <- c(2, 1)
Sigma <- matrix(c(1, 0.5, 0.5, 1), ncol = 2)
data <- expand.grid(seq(0, 4, length.out = 32),
                    seq(0, 2, length.out = 32))
data["dvalues"] <- dmvnorm(data, mean = mu,
                           sigma = Sigma)
```

Example ii

```r
library(tidyverse)
ggplot(data, aes(Var1, Var2)) +
  geom_contour(aes(z = dvalues))  +
  coord_fixed(ratio = 1)
```

Example iii



Example iv

```r
k <- 0.12
const <- -2*log(k*2*pi*sqrt(det(Sigma)))

# Generate a circle
# First create a circle of radius const
theta_vect <- seq(0, 2*pi, length.out = 100)
circle <- const * cbind(cos(theta_vect),
                        sin(theta_vect))
```

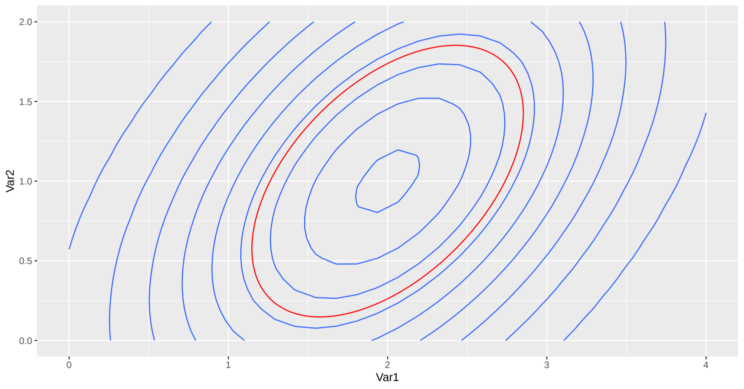
Example v

```r
# Compute inverse Cholesky
transf_mat <- solve(chol(solve(Sigma)))
# Then turn circle into ellipse
ellipse <- circle %*% t(transf_mat)
# Then translate
ellipse <- t(apply(ellipse, 1, function(row) row + mu)
```

Example vi

```r
# Add ellipse to previous plot
ggplot(data, aes(Var1, Var2)) +
  geom_contour(aes(z = dvalues)) +
  geom_polygon(data = data.frame(ellipse),
               aes(X1, X2), colour = 'red', fill = NA) +
  coord_fixed(ratio = 1)
```

Example vii



Elliptical distributions

- Elliptical distributions are a generalization of the multivariate normal distribution that retain the property that lines of constant density are ellipses.
- More formally, let $\mu \in \mathbb{R}^p$ and Λ be a $p \times p$ positive-definite matrix. If $\mathbf{Y}$ has density

$$f(\mathbf{Y}) = |\Lambda|^{-1/2} g\left((\mathbf{Y} - \mu)^T \Lambda^{-1} (\mathbf{Y} - \mu)\right),$$

  where $g : [0, \infty) \to [0, \infty)$ does not depend on μ, Λ, we say that $\mathbf{Y}$ follows an **elliptical distribution** with location-scale parameters μ, Λ, and we write $\mathbf{Y} \sim E_p(\mu, \Lambda)$.

Properties i

- The class $E_p(\mu, \Lambda)$ contains the multivariate normal distribution $N_p(\mu, \Lambda)$.
    - With $g(t) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}t\right)$.
- Affine transformations of elliptical distributions are again elliptical:
    - If $\mathbf{Y} \sim E_p(\mu, \Lambda)$ and B is invertible, then $B\mathbf{Y} + b \sim E_p(B\mu + b, B\Lambda B^T)$.
- We call $E_p(0, I_p)$ the class of *spherical distributions*.

Properties ii

- If $\mathbf{Y} \sim E_p(\mu, \Lambda)$, then its characteristic function is given by
$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \exp(i\mathbf{t}^T\mu)\psi(\mathbf{t}^T\Lambda\mathbf{t}),$$
for some real-valued function ψ.
- If $\mathbf{Y} \sim E_p(\mu, \Lambda)$ has moments of order 2, then $E(\mathbf{Y}) = \mu$ and $\mathrm{Cov}(\mathbf{Y}) = \alpha\Lambda$, where $\alpha = -2\psi'(0)$.

Proposition

Let $\mathbf{Y} \sim E_p(\mu, \Lambda)$, and write

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$
$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}.$$

Then $\mathbf{Y}_1 \sim E_{p_1}(\mu_1, \Lambda_{11})$ and $\mathbf{Y}_2 \sim E_{p_1}(\mu_2, \Lambda_{22})$.

Theorem

Let $\mathbf{Y} \sim E_p(\mu, \Lambda)$, and assume the same partition of μ and Λ as previously. Then

$$\mathbf{Y}_1 \mid \mathbf{Y}_2 = \mathbf{y}_2 \sim E_{p_1}(\mu_{1|2}, \Lambda_{1|2}),$$

where

$$\mu_{1|2} = \mu_1 + \Lambda_{12}\Lambda_{22}^{-1}(\mathbf{y}_2 - \mu_2),$$
$$\Lambda_{1|2} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}.$$

Unlike the normal distribution, the conditional covariance $\mathrm{Cov}(\mathbf{Y}_1 \mid \mathbf{Y}_2 = \mathbf{y}_2)$ will in general depend on $\mathbf{y}_2$.

First Example–Mixture of standard normal

- Let $\mathbf{Z} \sim N_p(0, I_p)$ and $w \sim F$, where F is supported on $[0, \infty)$.
- If we set $\mathbf{Y} = W^{1/2}\mathbf{Z}$, then $\mathbf{Y} \sim E_p(0, I_p)$ has a spherical distribution.
- Examples:
  - $P(W = \sigma^2) = 1$ gives the multivariate normal $N_p(0, \sigma^2 I_p)$.
  - $P(W = 1) = 1 - \epsilon$ and $P(W = \sigma^2) = \epsilon$ gives the *symmetric contaminated normal distribution*.
- We can generate data from these distributions by sampling W and $\mathbf{Z}$ *independently* and then calculating $\mathbf{Y}$.

Simulating data i

```r
set.seed(7200)

n <- 1000
p <- 2
Z <- rmvnorm(n, sigma = diag(p))
```

Simulating data ii

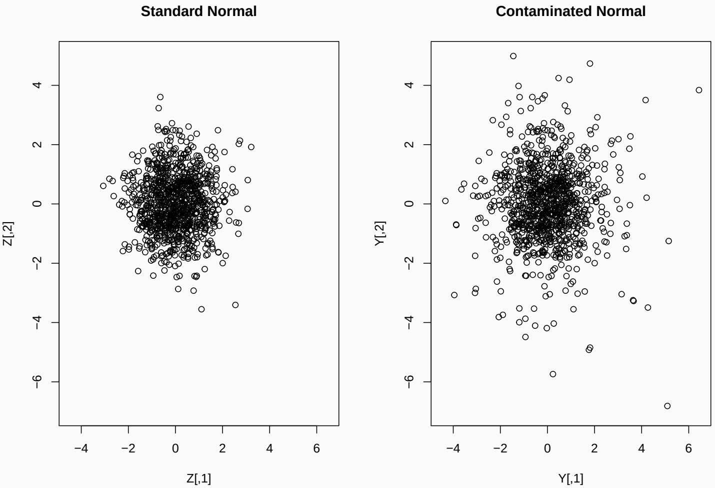
```r
sigma <- 2
epsilon <- 0.25
w <- sample(c(sigma, 1), size = n, replace = TRUE,
            prob = c(epsilon, 1 - epsilon))

Y <- w*Z
```

Simulating data iii

```r
# Plot the results
par(mfrow = c(1, 2))
plot(Z, main = 'Standard Normal',
     xlim = c(-4.5, 6.5), ylim = c(-7, 5))
plot(Y, main = 'Contaminated Normal',
     xlim = c(-4.5, 6.5), ylim = c(-7, 5))
```

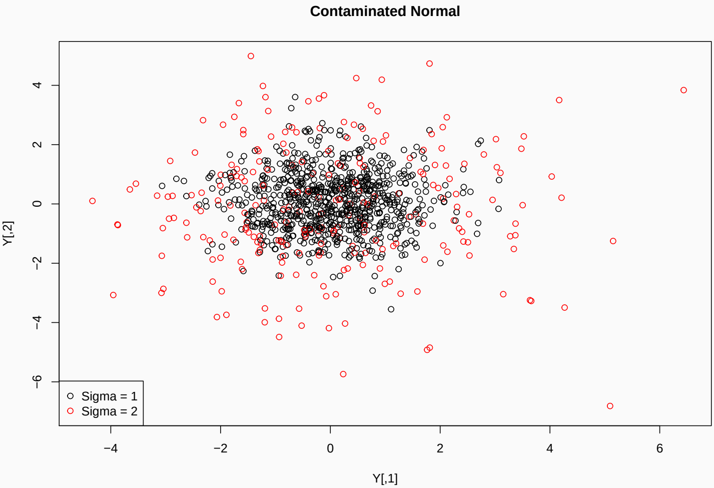
Simulating data iv



Simulating data v

```
# Colour points of Y according to
# which distribution they come from
plot(Y, main = 'Contaminated Normal', col = w,
     xlim = c(-4.5, 6.5), ylim = c(-7, 5))
legend("bottomleft", legend = c("Sigma = 1", "Sigma = 
       col = c(1, 2), pch = 1)
```

Simulating data vi



Simulating data vii

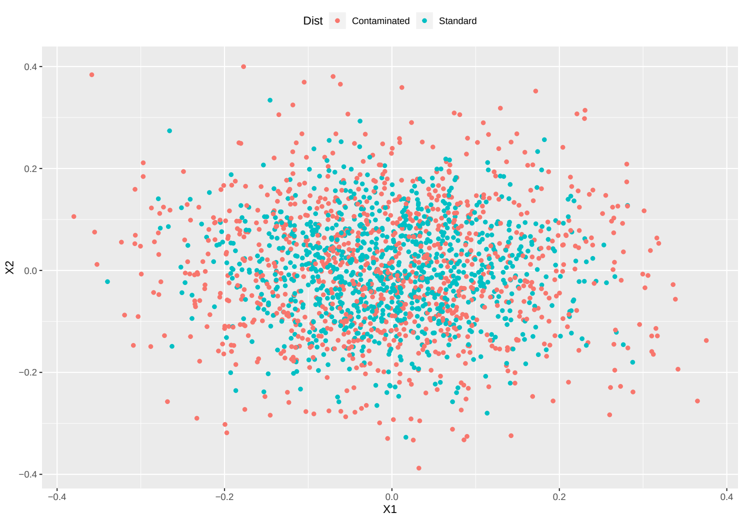
```r
# Let's look at the distribution of the sample means
B <- 1000; n <- 100
data <- purrr::map_df(seq_len(B), function(b) {
  Z <- rmvnorm(n, sigma = diag(p))
  Y <- sample(c(sigma, 1), size = n, replace = TRUE,
              prob = c(epsilon, 1 - epsilon)) * Z

  out <- data.frame(rbind(colMeans(Z), colMeans(Y)))
  out$Dist <- c("Standard", "Contaminated")
  return(out)
  })
```

Simulating data viii

```r
ggplot(data, aes(X1, X2)) +
  geom_point(aes(colour = Dist)) +
  theme(legend.position = 'top')
```

Simulating data ix



Second Example–t distribution i

- Let $\nu > 0$. If we take W in the mixture distribution above to be such that $\nu W^{-1} \sim \chi^2(\nu)$, we get the multivariate t distribution $t_{p,\nu}$. Its density is given by

$$f(\mathbf{Y}) = c_{p,\nu}(1 + \mathbf{Y}^T\mathbf{Y}/\nu)^{-(\nu+p)/2},$$

  where

$$c_{p,\nu} = \frac{(\nu\pi)^{-p/2}\Gamma\left(\frac{1}{2}(\nu+p)\right)}{\Gamma\left(\frac{1}{2}\nu\right)}.$$

Second Example–t distribution ii

- By relocating and rescaling, we can obtain the general multivariate t distribution $t_{p,\nu}(\mu, \Lambda)$: assume $\mathbf{Z} \sim t_{p,\nu}$ and set $\mathbf{Y} = \Lambda^{1/2}\mathbf{Z} + \mu$. The density of $\mathbf{Y}$ is now

$$f(\mathbf{Y}) = c_{p,\nu}|\Lambda|^{-1/2}(1 + (\mathbf{Y} - \mu)^T\Lambda^{-1}(\mathbf{Y} - \mu)/\nu)^{-(\nu+p)/2}.$$

- Note that the multivariate $t_{p,1}$ with $\nu = 1$ is known as the *multivariate Cauchy distribution*.

Second Example–t distribution iii

- The following side-by-side comparison may be helpful: Let $\mathbf{Z} \sim N(0, I_p)$, $\nu > 0$, $\mu \in \mathbb{R}^p$ and Λ $p \times p$ and positive definite.
  - $\mu + \Lambda^{1/2}\mathbf{Z} \sim N_p(\mu, \Lambda)$;
  - $\mu + \sqrt{W}\Lambda^{1/2}\mathbf{Z} \sim t_{p,\nu}(\mu, \Lambda)$, where $\nu W^{-1} \sim \chi^2(\nu)$.
- Finally, note that if $\mathbf{Y} \sim t_{p,\nu}(\mu, \Lambda)$, we have
  - $E(\mathbf{Y}) = \mu$, assuming $\nu > 1$;
  - $\mathrm{Cov}\,(\mathbf{Y}) = \frac{\nu}{\nu-2}\Lambda$, assuming $\nu > 2$.

Example i

```r
library(mvtnorm)
n <- 1000
mu <- c(1, 2)
Sigma <- matrix(c(1, 0.5, 0.5, 1), ncol = 2)
# Recall the multivariate case
Y_norm <- rmvnorm(n, mean = mu, sigma = Sigma)

colMeans(Y_norm)
```

```
## [1] 1.010602 1.990707
```

Example ii

```r
cov(Y_norm)
```

```
##             [,1]      [,2]
## [1,] 0.9937129 0.5059292
## [2,] 0.5059292 0.9983259
```

```r
# Now the t distribution
nu <- 4
Y_t <- rmvt(n, sigma = Sigma, df = nu, delta = mu)

colMeans(Y_t)
```

Example iii

```
## [1] 1.054561 2.041417
```

```
cov(Y_t)
```

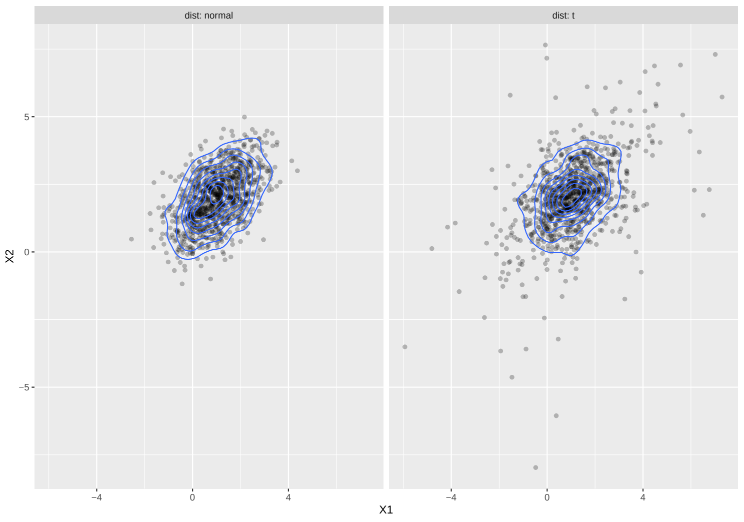
```
##              [,1]      [,2]
## [1,] 1.8399561 0.9004441
## [2,] 0.9004441 1.9164044
```

Example iv

```r
data_plot <- rbind(
  mutate(data.frame(Y_norm), dist = "normal"),
  mutate(data.frame(Y_t), dist = "t")
)

ggplot(data_plot, aes(X1, X2)) +
  geom_point(alpha = 0.25) +
  geom_density_2d() +
  facet_grid(~ dist, labeller = label_both)
```

Example v



Estimation i

- Given a random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from an elliptical distribution $E_p(\mu, \Lambda)$, we are interested in estimating μ and Λ.
- Recall that the sample mean and the sample covariance are *still* consistent:

$$\bar{\mathbf{Y}} \to \mu$$
$$S_n \to \alpha\Lambda.$$

- However, in general, they are no longer efficient.
  - You can build estimators with smaller variance.

Estimation ii

- The log-likelihood for our random sample is

$$\ell(\mu, \Lambda) = \sum_{i=1}^{n} \log \left(g \left((\mathbf{Y}_i - \mu)^T \Lambda^{-1} (\mathbf{Y}_i - \mu) \right) \right) - \frac{n}{2} \log |\Lambda|.$$

- Differentiating with respect to μ and Λ and setting the derivatives equal to zero, we get a system of equations:

$$\sum_{i=1}^{n} u(s_i) \Lambda^{-1} (\mathbf{Y}_i - \mu) = 0$$

$$\frac{1}{2} \sum_{i=1}^{n} u(s_i) \Lambda^{-1} (\mathbf{Y}_i - \mu)(\mathbf{Y}_i - \mu)^T \Lambda^{-1} - \frac{n}{2} \Lambda^{-1} = 0,$$

Estimation iii

where

$$u(s) = -2g'(s)/g(s),$$
$$s_i = (\mathbf{Y}_i - \mu)^T \Lambda^{-1} (\mathbf{Y}_i - \mu).$$

- Therefore, the MLE estimators (if they exist!) satisfy the following equations:

$$\hat{\mu} = \frac{\frac{1}{n}\sum_{i=1}^n u(s_i)\mathbf{Y}_i}{\frac{1}{n}\sum_{i=1}^n u(s_i)},$$
$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^n u(s_i)(\mathbf{Y}_i - \hat{\mu})(\mathbf{Y}_i - \hat{\mu})^T.$$

Estimation iv

- In other words, the MLE are in general **weighted** sample estimators.

Additional comments

- The MLEs do not have a closed form solution.
  - They must be computed using an iterative scheme.
- The existence and uniqueness of a solution to these estimating equations is a difficult theoretical problem.
- Alternatively, one can use *robust* estimators that have good properties for most elliptical distributions.
  - E.g M-estimators and S-estimators.
  - For details, see Chapter 13 of *Theory of Multivariate Statistics*
- On the Bayesian side of estimation, there is in general no closed form for the posterior distribution.
  - But efficient MCMC strategies can be developed for elliptical distributions.